FINDING COMMON GROUND IN MANAGING DATA USED FOR REGIONAL ENVIRONMENTAL ASSESSMENTS

STEPHEN S. HALE[1], LOWELL H. BAHNER[2] and JOHN F. PAUL[1]
[1] *U.S. Environmental Protection Agency, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI, 02882 USA;* [2] *NOAA Chesapeake Bay Office, 410 Severn Ave, Suite 107, Annapolis, MD, USA*

Abstract. Evaluating the overall environmental health of a region invariably involves using databases from multiple organizations. Several approaches to deal with the related technological and sociological issues have been used by various programs. Flexible data systems are required to deal with rapid changes in technology, the social and political climate for sharing and integrating data, and expectations of diverse users. Here we describe how the Environmental Monitoring and Assessment Program and the Chesapeake Bay Program manage their data for regional studies. These programs, which encompass areas of different geographic scales but face similar issues, have adopted some solutions in common, but also have tried some unique solutions suited to their needs. Understanding the tribulations and successes of these programs may help others attempting similar assessments. Both these programs have embraced distributed data systems that are managed by the organizations owning them. Both use common guidelines and policies that assure consistency and quality of data and information. These principles and tools comprise a flexible, sustainable approach that meets modern challenges of data management.

1. Introduction

Many environmental managers need to know the condition of environmental resources in regional (biogeographic, political, watershed) areas. They wish to know where in the region any problems are, how they can remedy them, and what they can do to prevent future problems. Managers would like to measure the effectiveness of management programs and practices. They require consistent information of known reliability that is indifferent to political, geographic, or watershed boundaries, but can be compiled in various ways.

In many regions, and particularly in the U.S. mid-Atlantic area, data that are useful for environmental assessments are being collected by several monitoring programs and individual research projects. These efforts are usually limited in time, space, and scope and are driven by different objectives. Taken together, their data are often spotty and involve multiple owners, software, hardware, standards, formats, degree of automation, and levels of documentation. Furthermore, data quality often changes abruptly at political or watershed boundaries. The consistent, comprehensive database that environmental managers wish they had, one that provides the data they need to answer their regional-scale questions, rarely exists. Seldom is there a political structure or funding to support the cost of develop-

Environmental Monitoring and Assessment **63:** 143–157, 2000.

ing a multidisciplinary data system that can be sustained indefinitely for a region the size of the U.S. mid-Atlantic.

Given this need and situation, what data management measures can help? Can data from existing programs be combined coherently enough to answer managers' questions? If not, could data from these programs be supplemented with new field work to help such assessments? What kind of data system could support such a task? Can it be simple enough to be sustained? In addition to technological issues of data sharing and integration, can programs overcome sociological (NRC 1995) and legal, economic, and policy (NRC 1997) factors that influence access to data? A number of programs have been developed to study regional environments and to address questions like these. For example, EPA sponsors studies by its ten Regional Offices, its regional programs (e.g., Chesapeake Bay Program, Gulf of Mexico Program), and regional studies of the Environmental Monitoring and Assessment Program (EMAP), such as the Mid-Atlantic Integrated Assessment (MAIA) and the Western Geographic Study. Interagency programs like the Committee on Environment and Natural Resources (CENR 1997) and the National Assessment of the US Global Change Research Program (e.g., the Mid-Atlantic Regional Assessment, USGCRP 1999) are grappling with the same issues. This paper illustrates some of these issues by describing how EMAP's MAIA and the Chesapeake Bay Program (CBP) responded to these challenges.

EMAP is a national program that was developed because disparate information from across the country made it difficult to construct a coherent national picture that would show the progress being made toward improving the environment (USEPA 1997). EMAP participates in regional studies to provide consistent information for biogeopolitical regions. The first such study, MAIA, was designed to develop approaches and tools for collecting and interpreting regional environmental data (Holland and DeMoss 1996). It is a pilot study to test the premises of the CENR (1997) national environmental monitoring framework. MAIA is a joint program of EPA's Office of Research and Development, EPA Region III, and collaborators. It includes EPA Region III (PA, DE, MD, WV, DC, and VA) and the watersheds of Chesapeake Bay, Delaware Bay, and Albermarle-Pamlico Sound (Figure 1). The MAIA region contains several existing geographic initiatives, such as the Chesapeake Bay Program, the Canaan Valley Institute, the Delaware River Basin Commission, and three National Estuary Programs. Each initiative would like to exchange data with their neighbors and participate in larger-scale activities. MAIA's data come from these and other monitoring programs and from its own supplemental sampling. For MAIA to succeed, it must be able to access, integrate, and analyze data collected by several programs.

The Chesapeake Bay Program has a Congressional mandate to restore the Chesapeake Bay (Figure 1) and protect its living resources. Begun in 1978 as a research and monitoring program, the CBP has evolved into a program shared by Virginia, Maryland, Pennsylvania, the District of Columbia, the Chesapeake Bay Commission, and the Federal government (CBP 1999). CBP's challenge is to stop the long-term decline in the health of the bay, restore its resources, and chart a course to prevent the increasing human population in the 166,000 km^2 basin from triggering a new decline in the bay. The problems and solutions for Chesapeake Bay are complex enough that over 50 agencies and organizations are working together through subcommittees to manage it. For years, raw data had been managed at a data center. Now, the Internet makes documents, graphs, maps, environmental indicators, and interpretations of data widely available. During 1996 and 1997, partner agencies created their own web sites, which helped make CBP information more accessible. On the other hand, creating too many web sites for the Chesapeake region could have hindered integrated assessments because procedures and

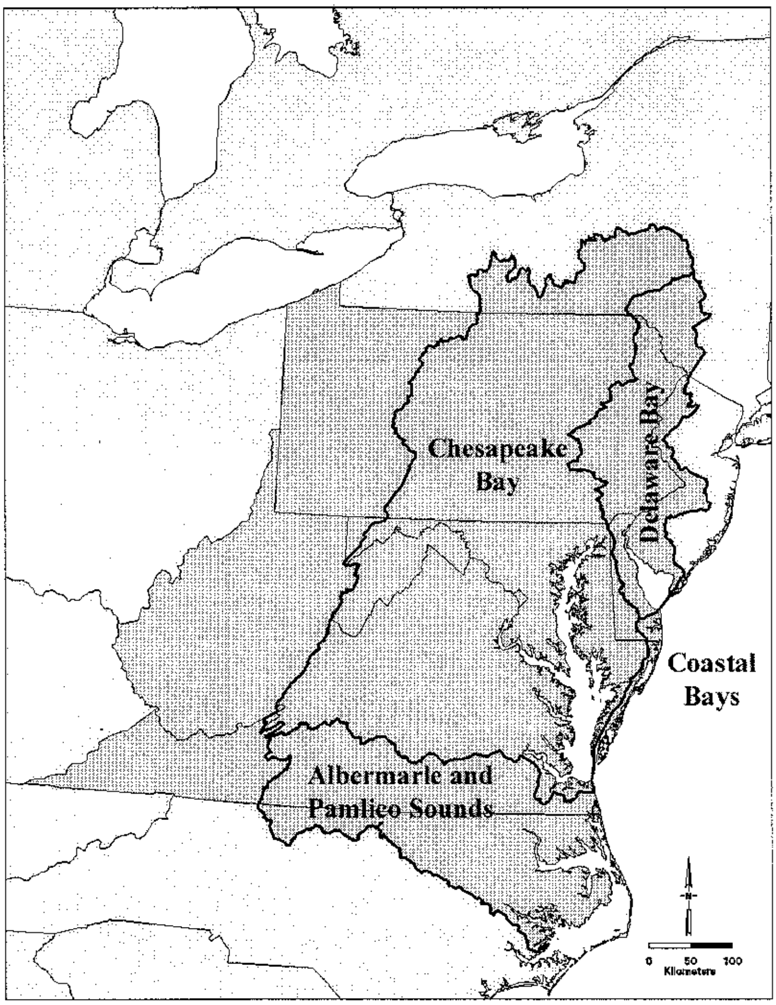

Figure 1. *The region of the Mid-Atlantic Integrated Assessment and the Chesapeake Bay Program, showing the major watersheds.*

tools for acquiring and analyzing information from distributed sources did not yet exist. A new vision was required if data managers were to successfully share information among so many organizations. A recent CBP Information Management workshop recommended that the CBP develop the Chesapeake Information Management System (CIMS) that would use standardized metadata, have common data standards, put as much information as feasible on the Internet, offer various data formats for diverse users, allow spatial and textual queries, and be easy for all to use.

This paper discusses how two regional studies share and integrate data from different data systems. The differences between MAIA and CBP give an idea of the range of possible approaches to the same problems. The CBP has a long-term commitment to the Chesapeake Bay and its watershed. Its program office has existed long enough for it to establish long-term monitoring programs with individual states. This has enabled CBP to develop more elaborate data management structures than programs like MAIA can develop. MAIA covers a larger area than CBP and includes multiple political units and watersheds. MAIA is not intended to be a long-term data-collecting program like the CBP. Instead, it relies on data collected independently by existing organizations. MAIA's structure requires a different kind of data system—less structured, more flexible, and more dependent on the voluntary participation of data-collecting organizations. MAIA's and CBP's approaches to data management help lead toward success of the CENR monitoring framework and could be useful to other regional studies.

2. EMAP Information Management

2.1 EMAP'S GENERAL APPROACH

No single organization can collect and maintain all the multidisciplinary data needed for assessing entire regions, but many organizations collect and manage pieces of the puzzle. EMAP has adopted the approach advocated by the US Global Change Research Program (USGCRP 1998) and the National Research Council (NRC 1997), where rather than trying to combine data into single, all-inclusive databases, programs are encouraged to manage their own databases in accordance with some common standards and formats. This allows data to be found and combined for analyses.

The two cornerstones of EMAP's approach are that: 1) original data are best managed, described, and maintained by the organization that collected them, and 2) the needs of researchers are best served by straightforward data-retrieval systems that allow them to download data to their own analytical software (Hale et al. 1999). EMAP uses a federated system of distributed databases with a core set of common standards and procedures. EMAP strives for simple ways to find data, to retrieve them, and to understand them. For this kind of data system to be sustainable, it must be kept simple and must focus on the primary goal of delivering high-quality data and metadata.

EMAP's data collection centers operate field computers, track samples, verify and validate raw data, process the initial data, operate analytical databases, write metadata, and provide long-term archival and distribution for low-level data. They send summary data and metadata to EMAP's web site for access by a broad audience (Figure 2). EMAP's information management coordinating office promotes common data policies, guidelines, and standards (e.g., FGDC 1998, ITIS 1999); operates internal and public web sites; serves data and metadata to secondary users; maintains the EMAP Data Directory; maintains the EMAP Bibliography; provides software and training; archives common data; and coordinates with other data management groups and systems (Hale et al. 1998).

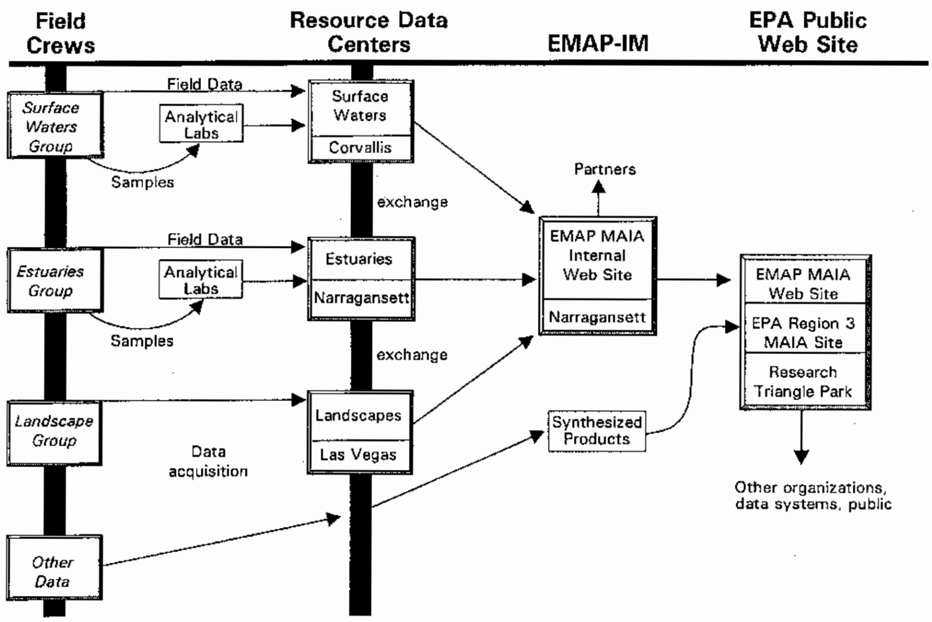

Figure 2. Data flow in the Mid-Atlantic Integrated Assessment.

2.2 MAIA INFORMATION MANAGEMENT

The objectives of MAIA information management are to:

- support analyses designed to answer the projects' assessment questions;
- produce state of the environment reports for Estuaries (USEPA 1998), Surface Waters (USEPA in prep.), and Landscape Ecology (Jones et al. 1997);
- promote continuously updated information systems that can support continuing environmental assessments as new data become available;
- promote full and open data sharing;
- provide long-term archiving for data from the supplemental sampling;
- encourage data sources to move toward common policies and standards for their data;
- promote ways of managing data that can be sustained by organizations in the region; and
- provide synthesized data products and results for the general public.

MAIA's Approach

MAIA takes a collaborative approach that uses partners' data more than earlier EMAP research did. Previously, EMAP controlled all aspects of managing data, from collection through quality assurance, processing, and analyzing, to distributing and archiving. MAIA's participating organizations manage data in their own ways. Because they will continue to collect data after the main EMAP work has finished, MAIA builds on existing systems in the region. For example, the ongoing CBP monitoring program was essential to the success of the MAIA estuaries study. The sociological aspects of these collaborative studies can be as important as their technical aspects (Farrey et al. 1999). A case study which documented data management practices of six environmental programs (NRC 1995) listed three sociological items among their ten keys to success: "Use a collaborative approach; account for human behavior and motivation; consider needs of participants as well as users." Although computer hardware, software, and networks are continuously improving, MAIA needs to pay careful attention to human interactions if it is to share data most effectively. Sharing data does not work well if based solely on governmental directives or on altruism. Partners share data because of mutual benefits from creating consistent data on scales larger than their own. MAIA tries to help participants identify and focus on these mutual benefits.

Shared data must be integrated for analyses. Three groups within EMAP (Landscape Ecology, Surface Waters, and Estuaries) collected data to supplement the ongoing monitoring by other organizations. In accordance with EMAP's core standards, each group's data management system is consistent for that resource. Each of the state of the environment reports produced by these groups combined data from ongoing programs and supplemental sampling. An additional benefit of exchanging data is that it promotes good data management (NRC 1995). In this case, the Landscape Ecology group needed data from streams to develop landscape indicators, and the Estuaries group needed landscape indicator data to help explain the variance of estuarine parameters. Because the groups interacted, they took a greater interest in common data standards and formats.

MAIA's Information Systems

Data from EMAP's resource groups and most of the other activities are already in data systems. It was not practical to develop a new, comprehensive database with a common structure (same tables and relationships) and dictionary (same names and codes for the parameters) because the size and cost of this kind of effort were beyond MAIA's resources. Nevertheless, MAIA needs a process to help bring data together for analysis. The process consists of initiating conventions, creating data assets, and then using these assets in syntheses. MAIA is an experimental program built on partnerships and goodwill. Because MAIA is an innovative program, it must manage its data innovatively, but must also keep the solutions simple and inexpensive so that they can be sustained by organizations in the region. MAIA adopted a hybrid approach that uses a variety of data management applications including: an inventory of monitoring programs, a data directory, analytical and summary databases with code translators to blend disparate conventions, metadata, and web pages.

The **Inventory of Environmental Monitoring Programs** (Jackson and Gant 1998) is a database that holds georeferenced information about data collection activities of the organizations in the region (NEMI 1999). This is a pilot project for the Committee on Environment and Natural Resources. The **EMAP Data Directory** is an Oracle database used for tracking EMAP datasets; datasets of interest can be found by searching it with a web browser (EMAP 1999).

Analytical databases were developed for individual resource groups or scientific disciplines. For example, the analytical database for MAIA Estuaries (Buffum and Hale 1998) is a consistent, region-wide SAS database that contains data collected by two EPA labs, the Chesapeake Bay Program and their state cooperators, two NOAA offices, the Delaware River Basin Commission, and the National Park Service. MAIA Estuaries is a multiagency partnership that agreed on a core set of measurements. The analytical database has a common structure and data dictionary that allow researchers from all the partners to use the same, consistent data in their regional analyses. These kinds of multiorganizational databases can be established more easily (i.e., common structures can be agreed upon) by limiting their duration, geographic extent, or type of data. In this case, the scope was limited to two years of data from estuaries within MAIA.

Metadata, which are essential for sharing and integrating data, are provided in EMAP's Data Catalog (and Data Directory). Metadata for regional partnership studies (e.g., MAIA Estuaries) are more complicated than for cases where the field work is done by a single entity. Even though all the partners in MAIA Estuaries adopted a core set of measurements and protocols that would produce comparable datasets, they each had slightly different ways of doing them. They all had their own field computer systems or paper forms, field method manuals, training programs, analytical programs, reporting requirements, database management systems, and web sites. Benthic community samples, sediment chemistry samples, and others were sent to different analytical laboratories for processing. The metadata had to track each datum and the steps from collecting through processing. Preparing the metadata needed at least as much effort as entering the data into the database. This is not uncommon (Michener 1998) and should be considered an inherent and essential cost of studies that collect data under the CENR framework.

MAIA's web pages, all physically located on EPA's public web server (*http://www.epa.gov)*, consist of three linked logical collections. The first collection is the MAIA pages on EMAP's web site (EMAP 1999), which holds **analytical and summary databases** of data collected by ORD's Landscape, Streams, and Estuaries groups. The site includes metadata, the Data Directory, reports, the EMAP Bibliography, contacts, and information on projects. This site is primarily for researchers. The second collection is the MAIA pages (MAIA 1999), which will be easier to use and is intended to provide information to managers and the public. There will be more graphics, more interpretation of results, and more discussion of what the results mean and what can be done about them. This site will include synthesized products that result from integrating data collected by EMAP with data from other monitoring programs. A data clearinghouse and warehouse are being considered to help link the geographic initiatives in the MAIA area and to store synthesized data products that may not have a home elsewhere. The third collection is the National Environmental Monitoring Initiative (NEMI 1999) site, which holds the Inventory of Monitoring Programs. All three of MAIA's web collections will provide the option to pose map-based queries and to obtain map outputs.

3. Chesapeake Bay Program Information Management

3.1 CBP'S GENERAL APPROACH

The CBP Data Center has been one of the largest repositories of data on Chesapeake Bay. Water quality measurements (primarily temperature, salinity, and dissolved oxygen) are available from 1906 on, although they are sparse until 1979, when the bay-wide monitoring program began. Since then, extensive data have been added for water quality, living resources, and other parameters. These data were managed as SAS datasets on a single minicomputer to provide a common platform for analysis and integration. Data were submitted periodically by state and academic organizations. A central staff standardized the names of variables and stations, corrected errors reported by users, and documented each dataset using a consistent format. This central repository was costly to maintain because data held in computers at the CBP and state agencies soon became de-synchronized as edits and updates were not always applied to all copies of the data. This central data center was critical to the success of the CBP during the pre-Internet years.

The advent of the Internet required new and streamlined processes for collecting, organizing, and distributing information. By allowing the generating organizations to own and keep their data, CBP hoped that pride of ownership would stimulate data owners to improve data quality and that the overall cost of managing information would decrease. To date, not only have there been significant cost savings from publishing CBP data and information over the Internet, but data are available sooner because of fewer layers handling the information.

CBP's web site provides the primary data used in programmatic assessments. For example, data on nutrients can be selected and displayed as interpolated bay-wide images of concentrations over time. Models for the watershed and the bay are run under various management scenarios to predict how the system would respond to reductions of nutrients from controlling point and nonpoint sources and from changing land use stemming from new regulations and new zoning. CBP is also trying to model how the bay's living resources will respond to projected improvements in water quality.

3.2 CHESAPEAKE INFORMATION MANAGEMENT SYSTEM

In October 1996, the Executive Council (State Governors, the Mayor of the District of Columbia, the Chairman of the Chesapeake Bay Commission, and the EPA Administrator) signed the "Strategy for Increasing Basin-wide Public Access to Chesapeake Bay Information" (CBP 1999). This document directs the Program to maintain a coordinated data management system that will provide timely information on the progress of the restoration programs and will help citizens understand Chesapeake Bay, its problems, the policies and programs designed to help it, and what they as citizens can do to maintain the bay's resources. The document requires all participating agencies to implement the basin-wide Chesapeake Information Management System (CIMS). This is a much more directed approach than exists in MAIA.

In response to this strategy, the CBP Information Management Subcommittee developed policies and guidelines for directing CIMS-related activities (CBP 1999). The document provides guidance to all participating organizations regarding locational accuracy, map coordinates, metadata, naming stations, setting up dictionaries of common data, designing databases, dates, method codes, reporting data, and deliverables. These form the basic rules for creating and maintaining coordinated, distributed information systems. They were also designed to help other organizations standardize information that they would distribute over the Internet.

CIMS's success depends on each partner organization doing its share. These organizations sign agreements that publicly proclaim their long-term roles and responsibilities in maintaining distributed databases that incorporate common standards and documentation. Getting these agreements signed focuses the upper management of all organizations on their responsibilities for properly managing data and rapidly publishing information. Success also depends on the partners following the same plan. Common policies and guidelines become critical when merging data from different locations. This everyday occurrence for analyzing how the region's quality of air, land, and water affect its living resources requires that all partners use the same language; namely, consistent data dictionaries for every database. Now that the Internet has provided the means for sharing information, each organization must publish in the same language so that others can more easily use that information. This does not mean that all organizations must use the same software and hardware—it means only that their data structures must be consistent.

Distributed information networks hinge on their interfaces between users and data. Users should be able to select information of interest through a "smart user interface" that sends the search request to the metadata search engine, which acts like a library card catalog. It searches the metadata and identifies where the information is housed. The user's query is then forwarded to remote computers that have the requested information. Those computers extract the information and send it to the requesting computer, which packages it into a useful format and delivers it to the user's computer. The goal of this "CIMS at the desktop" approach is to provide useful data, graphs, maps, and documents to the user, with the capability of digging deeper for more detailed information. The user will be able to define a "Project" that will store all the retrieved information in a meaningful way. As "Projects" archive information related to a given subject area, they become sources of information that can be shared among workgroups so that there is no need to duplicate work already done. Metadata form the critical element for the next-generation search engines to locate information. They provide the information required to determine if the desired information exists, where it exists, how it is formatted, and how it can be accessed.

This approach represents a fundamental change in information handling and processing. It shifts the focus of managing technical data from databases to an Information Pyramid (Figure 3) of a range of types of information. At the bottom are the monitoring data that are required for technical evaluations and assessments. These data are primarily of interest to researchers and to analysts who evaluate data to see if program policies and criteria are being achieved. The analytical procedures, tools, models, and methods are important to those who need to assess the quality of information or how it was interpreted. In the past, intermediate graphs and interpretations were available only to the analyst, but now these products are understood to be useful to a wider audience and need to be shared. Included in this category are results of models, which may require months of calibration and test-

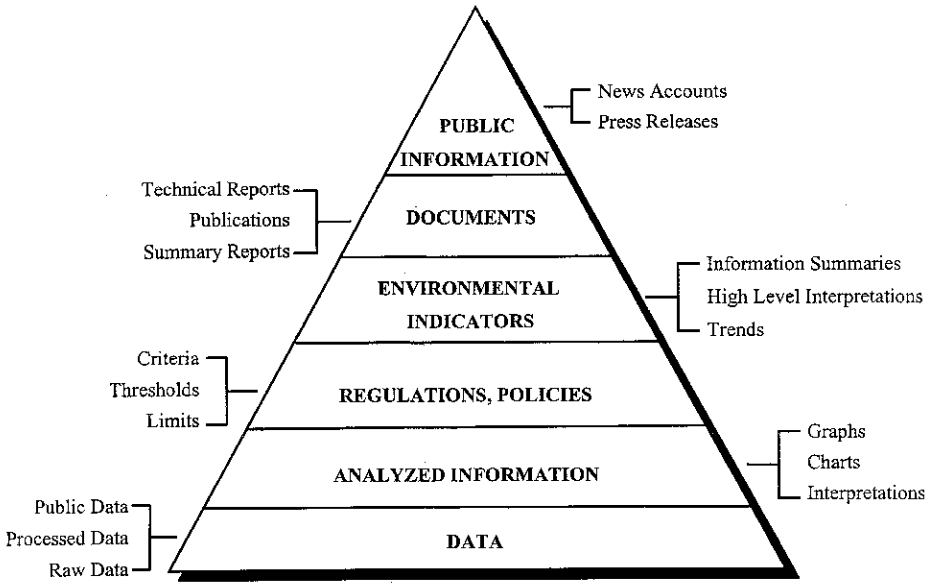

Figure 3. *The Chesapeake Information Management System's Information Pyramid shows the types of data, policies, analytical tools, and other information that are being managed for programmatic and public access.*

ing to develop even a single scenario. The "State of the Bay" report (CBP 1997) will become a living web document that delivers updated information to a broad audience. In contrast, MAIA serves the same levels of the information pyramid through its three collections of linked web pages.

Early designs of a smart user-interface took a bottom-up approach that used HTML requests to locate existing web pages. The first improvement was converting to dynamic pages that displayed data extracted from databases. The next-generation interface is being built on a Geographic Information System (GIS) platform that will overcome HTML's limited ability to deal with geographic information. A GIS can create maps and link them with data from any level of the Information Pyramid. The CBP, through use of GIS products, is developing a powerful system for finding and retrieving metadata and for displaying dynamic text, documents, graphics, and maps through the same interface. A prototype of this new vision for CIMS has shown how information and technology can rapidly answer environmental management questions. It also demonstrates the need for easily merging information from distributed sources. Now the challenge is to implement metadata servers that are intelligently linked to these distributed information sources.

4. Conclusions

To answer environmental questions about large regions, data are needed from multiple databases that are maintained by multiple organizations for their own purposes. These data may also be in multiple formats unless formal agreements, such as those of the CBP, can be established. In either case, data must be shared and integrated. Encouraging owners of data to move toward common standards, directories, and data descriptions for databases with distributed ownership makes it easier to find and deal with data of interest. In regional studies, sociological issues bearing on people's willingness to share data and to participate in integrated efforts may be as important as technological issues. Hardware, software, and networks are improving so quickly that the human interactions needed to share and integrate data may lag behind. Partners are motivated to share data when they can clearly identify mutual benefits.

Data systems must balance the need for integrating data within a single design against the cost and practicality of doing so. The CBP can work toward a consistent distributed database system because their Executive Council stated that this was a goal and all their partners have formally endorsed the idea. MAIA is not yet at that point, however, because it includes a much larger area with more diverse partners who tend to collaborate internally more than with MAIA as a whole. It is not likely that all the geographic initiatives in the area would adopt a common database design. Consistent database structures are not essential for regional environmental assessments—they simply ease the burden on the data analysts and other users. If these assessments are done frequently, a consistent structure should be considered. This is easier to accomplish if top-level managers from all the partners are convinced it is a good idea. Similar challenges have been faced by other large regional assessment programs (NRC 1997, USGCRP 1999). Success of the CENR monitoring framework depends on data sources agreeing upon certain core standards and being willing to make data and metadata available. Metadata are vital. The MAIA experience has shown that although compiling metadata from multiorganizational studies is not a trivial task, it must be considered an inherent cost of these kinds of studies where data may be used by someone years later for purposes unknown to the database designers.

Information management systems must always remain flexible enough to incorporate the new technologies that allow us to access more information faster and to share it more easily. A diversity of approaches may provide some resilience against adverse external forces such as vagaries in funding, fads in monitoring, and the comings and goings of programs. Varied approaches are needed to deal with the programs' many uses for data, levels of competence in managing data, and high rates of change. But the data management solutions must not be more complex than can be sustained by entities within the region. Above all, data management must ensure the integrity, quality, description, and long-term accessibility of the data.

Users are continually asking more of environmental information systems. Unless those systems keep pace with expectations, today's good idea may become tomorrow's problem. "User-friendly" systems carry a price, however. Like PC software, information systems that become more powerful and easier for the user also become more complex to develop and to maintain. Users must always remember that information systems are no better than their data. They must not allow an elegant interface to fool them into thinking otherwise, nor should they relax their healthy skepticism when looking at a striking GIS map that integrates disparate data. Further, a steady stream of new high-quality monitoring data are essential for these systems to retain their value.

MAIA and CBP share common ground in such things as distributed data systems, centralized metadata based on the FGDC format, species and chemical codes, web sites that serve all levels of the information pyramid, and web-based user interfaces that can access maps and relational databases. The technologies and procedures described in this paper have made information more readily available to scientists, managers, and the public in the U.S. mid-Atlantic area. Accessibility is a major goal of EPA's information management policy. As private corporations know, well-managed information is the most valuable asset of an enterprise. Developers of regional information systems who follow the NRC (1995, 1997) recommendations will have a good compass for guiding them toward success.

MAIA and CBP information systems will continue to use distributed databases operated by the data collectors and owners. Trends are toward centralized "intelligent" metadata that know what data exist, where to find them, and how to merge data from different sources. The CBP will use such metadata to fuel CIMS's next generation search engine. MAIA will do the same with a suite of directories, inventories, and centralized metadata. Systems will be more powerful and easier to use, with map-based interfaces where users can query from a map and have their output delivered as a data table, graph, or map. The Internet, web browsers that can access databases and maps, and the increasing ease of sharing data and information are the driving forces behind the dynamic changes in environmental information management today. Given continuous high-quality monitoring data, these developments will improve our ability to assess the condition of regional environments and thereby improve our ability to manage them.

Acknowledgments

L. Rossner, P. Bradley, and J. Rosen have contributed to the EMAP approach to data management for regional environmental assessments. A host of others has helped develop the EMAP information systems. R. Magnien, R. Batiuk, P. Hissom, M. Olson, R. Bahner, J. Johnson, and many other staff and committee members have contributed to the design and development of the Chesapeake Information Management System. K. Rahn, J. Kiddon, H. Walker, C. Strobel, and two anonymous reviewers suggested improvements to earlier versions of this article. This is contribution number NHEERL-NAR-2089 of the Atlantic Ecology Division.

References

Buffum, H. and Hale, S.: 1998, *MAIA-Estuaries 1997–1998. Data Format and Transfer Manual*, Atlantic Ecology Division, NHEERL, U.S. Environmental Protection Agency. Narragansett, RI, USA.

CBP: 1997, *State of the Bay*, Chesapeake Bay Program. Annapolis, MD, USA.

CBP: 1999, Chesapeake Bay Program web site, [*http://www.chesapeakebay.net*].

CENR: 1997, *Integrating the Nation's Environmental Monitoring and Research Networks and Programs: A Proposed Framework*, Committee on Environment and Natural Resources, National Science and Technology Council, Washington, DC, USA.

EMAP: 1999, Environmental Monitoring and Assessment Program web site, [*http://www.epa.gov/emap*].

Farrey, P.M., Mooney-Seus, M.L. and Tausig, H.C. (eds.): 1999, 'Out of the Fog: Furthering the Establishment of an Electronic Environmental Information Exchange for the Gulf of Maine,' New England Aquarium, *Aquatic Forum Series*, Report 99-1, Boston, MA, USA.

FGDC: 1998, *Content Standard for Digital Geospatial Metadata, version 2.0,* FGDC-STD-001-1998. Federal Geographic Data Committee, Washington, DC, USA.

Hale, S.S., Hughes, M.M, Paul, J.F., McAskill, R.S., Rego, S.A., Bender, D.R, Dodge, N.J., Richter, T.L and Copeland, J.L.: 1998, 'Managing Scientific Data: The EMAP Approach', *Environmental Monitoring and Assessment* **51**, 429–440.

Hale, S., Rosen, J., Scott, D., Paul, J. and Hughes, M.: 1999, *EMAP Information Management Plan: 1998–2001,* EPA/620/R-99/001, U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC, USA.

Holland, M.M. and DeMoss, T.B.: 1996, 'The Mid-Atlantic Integrated Assessment: Focus on Process', in: *North American Workshop on Monitoring for Ecological Assessments of Terrestrial and Aquatic Ecosystems*, Bravo, C.A. (ed.), US Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, USA. pp. 194–211.

ITIS: 1999, Integrated Taxonomic Information System web site, [*http://www.itis.usda.gov*].

Jackson, L.E. and Gant, M.P.: 1998, 'An Interactive, Spatial Inventory of Environmental Data in the Mid-Atlantic Region', *Environmental Monitoring and Assessment* **51**, 325–329.

Jones, K.B., Riiters, K.J., Wickham, J.D., Tankersley Jr., R.D., O'Neill, R.V., Chaloud, D.J, Smith, E.R. and Neale, A.C.: 1997, *An Ecological Assessment of the United States Mid-Atlantic Region: A Landscape Atlas,* EPA/600/R-97/130, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, USA.

MAIA: 1999, Mid-Atlantic Integrated Assessment web site [*http://www.epa.gov/maia*].

Michener, W.K.: 1998, 'Ecological Metadata', in: *Data and Information Management in the Ecological Sciences: A Resource Guide*, Michener, W.K., Porter, J.H. and Stafford, S.G. (eds.), LTER Network Office, Univ. of New Mexico, Albuquerque, NM, USA.

NEMI: 1999, National Environmental Monitoring Initiative web site, [*http://www.epa.gov/monitor*].

NRC: 1995, *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data*, National Research Council. National Academy Press, Washington, DC, USA.

NRC: 1997, *Bits of Power: Issues in Global Access to Scientific Data*, National Research Council, Washington, DC, USA.

USEPA: 1997, *EMAP Research Strategy*, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA.

USEPA: 1998, *Condition of the Mid-Atlantic Estuaries*, EPA/600/R-98/147, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA.

USEPA: *in prep.*, *Mid-Atlantic Highlands State of the Streams*, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA.

USGCRP: 1998, *Data Management for Global Change Research. Policy Statements for the National Assessment Program, July 1998*, U.S. Global Change Research Program, National Science Foundation, Washington, DC, USA.

USGCRP: 1999, Mid-Atlantic Regional Assessment of Climate Change Impacts web site. *[http://lumen.deasy.psu.edu/mara]*.